**PERFORMANCE MEASURES IN MACHINE LEARNING MODELS**

• Accuracy

• Weighted (Cost-Sensitive) Accuracy

• Lift

• Precision/Recall – F – Break Even Point

• ROC – ROC Area

CONFUSION MATRIX: To demonstrate the performance measures

|  |  |  |
| --- | --- | --- |
|  | Predicted 0 (NO) | Predicted 1(YES) |
| True 0 (NO) | a (Correct) | c (False Positive) |
| True 1 (YES) | b (False Negative) | d (Correct) |

Overall, how often is the classifier correct?

**ACCURACY** = (a+d)/(a+b+c+d)

Disadvantages:

1. Assumes equal cost for both types of errors
2. 99% accuracy may mean excellent, good, mediocre, poor or terrible model depending on the problem.

**LIFT**= [a/(a+b)]/[(a+c)/(a+b+c+d)]

not interested in accuracy on entire dataset • want accurate predictions for 5%, 10%, or 20% of dataset • don’t care about remaining 95%, 90%, 80%, resp. • typical application: marketing

Overall, how often is the classifier wrong?

**Misclassification rate**=(b+c)/(a+b+c+d)

When it predicts yes, how often is it correct?

**PRECISION** = d/(c+d)

When it's actually yes, how often does it predict yes?

**RECALL (True positive rate or Sensitivity)** = d/(b+d)

When it's actually no, how often does it predict yes?

**False positive rate**= c/(a+c)

When it's actually no, how often does it predict no?

**Specificity**= a/(a+c)

How often does the yes condition actually occur in our sample?

**Prevalence** = (b+d)/(a+b+c+d)

**BreakEvenPoint** is the point where Precision==Recall

**F1 Score** = 2\*((precision\*recall) / (precision + recall)).

It is also called the F Score or the F Measure. Put another way, the F1 score conveys the balance between the precision and the recall.

**ROC: Graph of sensitivity vs 1-specificity**

Sensitivity = Recall = a/(a+b) (Rate of correct 1 predictions)

Specificity = d/(d+c) (Rate of correct 0 predictions)

1-Specificity (Rate of incorrect 0 predictions: false positives)

• ROC Area:

– 1.0: perfect prediction

– 0.9: excellent prediction

– 0.8: good prediction

– 0.7: mediocre prediction

– 0.6: poor prediction

– 0.5: random prediction

–<0.5: something wrong!

Summary

• the measure you optimize to makes a difference

• the measure you report makes a difference

• use measure appropriate for problem/community

• accuracy often is not sufficient/appropriate

• ROC is gaining popularity in the ML community